Arizona's Measure of Academic Progress 2006

Summary and Technical Report

Tom Horne
Superintendent of Public Instruction
Arizona Department of Education

Research & Policy
Arizona Department of Education
Robert Franciosi, Deputy Associate Superintendent
Achieve@azed.gov

Published by the Arizona Department of Education, January 2007.

Introduction

This report presents the Measure of Academic Progress (MAP) analysis for the school year 2005-06. The MAP measures academic growth at the student level using Arizona's criterion-referenced AIMS test. This report presents average student progress in Arizona's public schools by grade and subject. It also provides supporting material to help interpret schools' MAP scores.

This report launches the new Measure of Academic Progress. Although the Arizona Department of Education has been presenting MAP analyses for seven years, previous years' MAP scores have out of necessity been calculated using the state's norm-referenced test. Starting with this report, MAP will directly measure student progress in learning state standards as measured by performance on the AIMS.

The new MAP uses a conventional value-added approach to measuring student growth. Value-added analysis is accepted by scholarly researchers as a valid method to evaluate the impact of programs and schools. In addition to using a different test and a new method of measuring growth, the new MAP is also the first time student records were matched across years using a unique student identifier (the SAIS ID).

Some History

The Measure of Academic Progress (MAP) is the state's original measure of school performance. It preceded AZ LEARNS, the state's current school accountability system, and is incorporated into the AZ LEARNS school profile. When the MAP was developed, the only test given statewide at all grades was the Stanford 9. The Stanford 9 was a norm-referenced test so it measured student performance against a national average instead of a state standard i.e. student's were told that they performed in the 35th or 75th percentile, not that they met or did not meet state standards.

The former MAP deemed a student to have made one-year's-growth (OYG) if that student held steady against the national average. Students were considered to have held steady if their score fell in the same stanine* or higher for two consecutive years. For example, if a student scored in the 3rd stanine in the fourth grade and scored in the 3rd or higher stanine in the fifth grade, that student was considered to have made one year's growth. Students who scored in the highest stanine (9) and fell to the next highest (8) in the following year were still considered to have made OYG. Students who remained in the lowest stanine (1) from one year to the next were not deemed to have made OYG. MAP scores for schools were found by calculating the percentage of students in a school making OYG in math and in reading.

* Stanine, short for "standard nine" is a method of reporting norm-referenced test results on a standard scale. Typically, a stanine is one of 9 divisions of the normal or bell curve. See http://www.mathnstuff.com/math/spoken/here/2class/90/stanine.htm for a very helpful diagram.

The AIMS Test

The new MAP measures student achievement using the Arizona's Instrument to Measure Standards Dual Purpose Assessment (AIMS). The AIMS is a criterion referenced test rather than a norm referenced test. That is, AIMS measures a student's success in meeting Arizona's academic standards rather than how that student performed compared to a national average. Consequently compared to the old MAP, the new MAP is a better measure of school effectiveness in teaching students what they should be learning, i.e. the state academic standards.

AIMS results are reported in four performance levels:

Exceeds the Standard – denotes superior academic performance on challenging subject matter reflected by the content standards

Meets the Standard – denotes solid academic performance and understanding of the state content standards

Approaches the Standard – denotes partial understanding of the skills and knowledge necessary for proficient work at grade level

Falls Far Below the Standard – denotes insufficient understanding of the prerequisite skills. Students who achieve at this level have serious gaps in knowledge and skills and may require remediation.

Performance levels of meets and exceeds are considered proficient.

The AIMS has been given to students since the 2000 school year, however all grades three through eight have been tested operationally only since the 2005 school year. In 2005 the test underwent a standard setting and a vertical scale was developed for grades three through eight. The scale ranges from 200 through 800, with successively higher cut scores for the three performance levels at each grade. Tables 1 and 2 show the cut scores for the performance levels.

Table 1. Mathematics Cut Scores for Grades 3 Through 8

	Grade					
	3	**4**	**5**	**6**	**7**	**8**
Approaches	386	414	442	463	484	505
Meets	420	448	476	496	517	537
Exceeds	492	521	550	574	599	623

Table 2. Reading Cut Scores for Grades 3 Through 8

	Grade					
	3	**4**	**5**	**6**	**7**	**8**
Approaches	379	402	424	433	443	452
Meets	431	450	468	478	489	499
Exceeds	516	536	556	571	587	602

Matching and Inclusion

ADE has the ability to track students through their career in Arizona's public schools from kindergarten through graduation via its Student Accountability Information System (SAIS). Since the 2001-2002 school year, students enrolling in Arizona schools have been assigned a unique identifier, the SAIS ID, that remains with them across grades, schools, and districts until they graduate from high school. Starting with the 2003-04 school year, SAIS IDs have been attached to testing data.

In previous years, matching for the MAP was based on student names, birthdates, and gender. ADE achieved a highly successful match rate via these means. The new MAP matches students using SAIS IDs. The table below shows the results from matching 2005 to 2004 test data using student SAIS IDs. The method achieves nearly a 90 percent match rate that does not differ by program membership or by proficiency level. As can be expected, the match rate is lower for mobile students and those not enrolled in a school for the full academic year.

Table 3. Match Rates

Subgroup	Match Rate
Total	89%
English language learners	85%
Free/Reduced lunch	89%
Special Education	89%
Full academic year	93%
Non-full academic year	63%
Proficient	94%
Not-proficient	91%

A student is included in the new MAP analysis if:

1. She took the AIMS test in the same subject area for two consecutive years;
2. She had a valid test score;
3. She did not take the test in either year with alternate accommodations;
4. Her scores were successfully matched from one year to the next.

Previously, the MAP analysis included students only if they were in the same school for two consecutive years, the reason being that a fair and valid evaluation of school performance should only be based on students the school had an ample opportunity to teach. This rule resulted in up to 45 percent of successfully matched students being excluded from the analysis. Also, this rule caused the grades evaluated to change from school to school due to different grade configurations.

It is desirable that as many students as possible are included in the MAP analysis for two reasons: schools should be held accountable for the achievement of *all* students; and large sample sizes provide more valid results. Standard statistical techniques can account for individual student characteristics such as mobility, and still produce a valid and fair

analysis. Consequently, the new MAP analysis includes all students with valid and comparable scores, with a special adjustment made for students not enrolled in a school for the full academic year. English language learners regardless of years in the program and special education students who have taken the AIMS without alternate accommodations are included in the MAP analysis.

Measuring Academic Growth

The table below shows average, student-level growth measured by change in AIMS scales scores (scale score 2006 – scale score 2005) from 2005 to 2006.

Table 4. Growth by Subject and Grade

Subject	Grade 2006	Number of Matched Scores	Average	Standard Deviation	Minimum	Maximum
Math	4	71,784	35	33	-165	248
	5	71,332	29	33	-179	248
	6	71,615	17	33	-196	224
	7	70,522	28	32	-178	238
	8	70,805	15	31	-200	273
Read	4	71,948	22	31	-155	240
	5	71,406	21	31	-180	253
	6	71,735	11	29	-182	237
	7	70,657	19	30	-174	250
	8	71,054	10	32	-200	267

The table below gives the descriptive statistics for student level growth (scale score 2006 – scale score 2005) across all grades and subjects by student performance level in 2005. The table shows a pronounced ceiling effect; that is, growth decreases the higher the starting point. Other evidence such as regression analysis and plots of the data also support this conclusion.

Table 5. Average Growth by Performance Level

AIMS Performance Level in 2005	Scale Score Growth from 2005 to 2006 Across All Grades and Subjects				
	Number	Average	Standard Deviation	Minimum	Maximum
FFB	109,270	51.8	63.1	-138	444
A	244,048	37.2	44.3	-239	273
M	603,807	20.4	37.0	-331	267
E	110,369	3.4	43.5	-470	213

The Growth Index

In order to control for the ceiling effect and for student mobility, the MAP analysis is done using a standard value-added model. The value-added model is used to calculate an estimate of expected growth for each student for each subject. The expected growth is then subtracted from actual growth to determine the student's Growth Index.

Example. Student A scored 478 on the 6th grade math test in 2005. Her expected growth for 7th grade in 2006 is 38. In 2006 she scores a 528 on the 7th grade math test. Her actual growth is 528 – 478 = 50. Her growth index is 50 – 38 = 12.

Example. Student B scored 490 on the 4th grade reading test in 2005. His expected growth for 5th grade in 2006 is 12. In 2006 he scores a 500 on the 5th grade reading test. His actual growth is 500 – 490 = 10. His growth index is 10 – 12 = -2.

The results of the MAP analysis for students and schools are reported in terms of the growth index. Growth indices for individual students can be averaged across schools, grades, and subjects to measure performance.

Calculating Expected Growth

The expected growth for an individual student is calculated using the following formula:

Expected Growth = A + B X (2005 scale score) + C X (FAY);

where FAY = 1 if the student has been enrolled for the full 2006 academic year and FAY = 0 if not. A student is considered to have been enrolled the full 2006 academic year if she has enrolled within the first two weeks of the school year and remained enrolled up to the time of the test.

The parameters A, B, and C are given in the tables below:

Table 6. Parameters for Calculating Expected Growth for Math

Student's Grade Current Year	A	B	C
4	98.9308	-0.1514	6.027
5	107.7715	-0.174	5.7754
6	75.6373	-0.1268	6.999
7	121.1295	-0.1896	7.1203
8	54.1785	-0.0823	7.0308

Table 7. Parameters for Calculating Expected Growth for Reading

Student's Grade Current Year	A	B	C
4	120.9638	-0.2269	3.7108
5	153.1619	-0.2879	4.106
6	105.8317	-0.2016	4.343
7	88.3119	-0.1492	5.1193
8	89.8856	-0.1688	7.0786

Example. Student A scored 478 on the 6th grade math test in 2005. She has been enrolled in her current school all year. Her expected growth for 7th grade in 2006 is 121.1295 + (-0.1896) X 478 + 7.1203 = 37.62 = 38.

Example. Student B scored 490 on the 4th grade reading test in 2005. He has not been enrolled in his current school for the full year. His expected growth for 5th grade in 2006 is 153.1619 + (-0.2879) X 490 = 12.09 = 12.

The numbers in the above tables are the estimated parameters resulting from the following value-added model:

$$Y_i = \alpha + \beta_1 Scalscor05_i + \beta_2 FAY_i + \varepsilon_i$$

The i's indicate individual students. Y_i is the student's academic growth as measured by the change in her AIMS scale score from one year to the next: $Y_i = Scalscor06_i - Scalscor05_i$. FAY_i is the indicator for full academic year described above, and ε_i is the conventional normally distributed random error term.

The equation was estimated by grade and subject using ordinary least squares. All parameter estimates were significant at the 99 percent level. The table below provides statistics for the goodness of fit. In general, the model does a better job of explaining differences in reading scores than math scores, although overall it explains very little of the difference in scores among students.

Table 8. Goodness of Fit Statistics for Value Added Regression

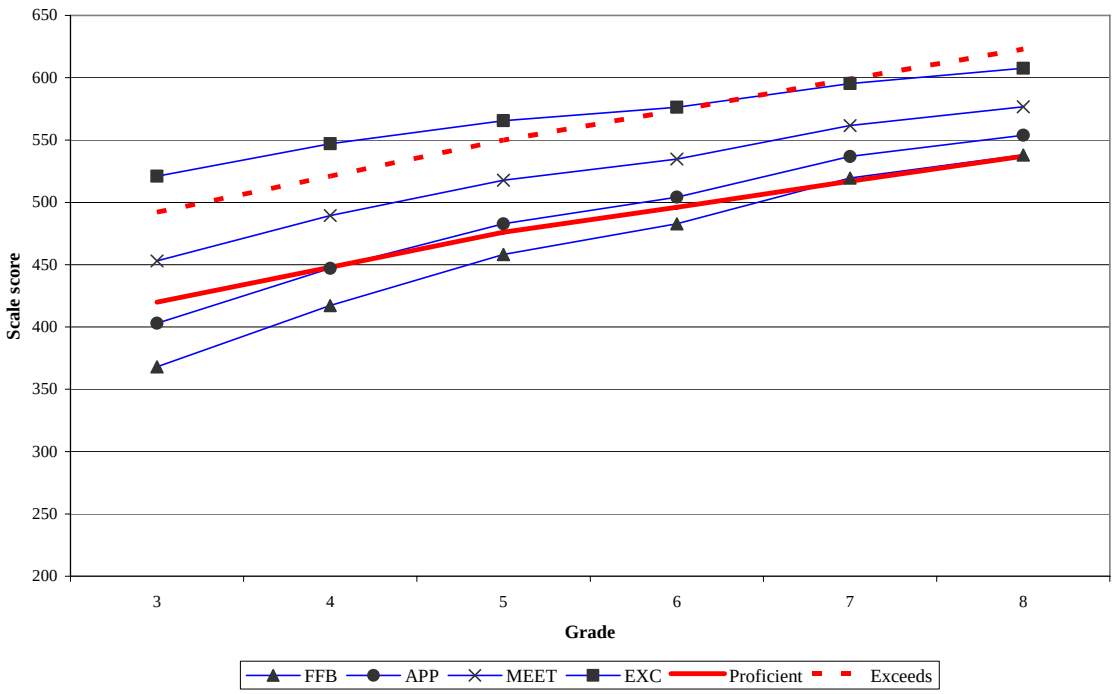
Subject	Grade2006	Root Mean Square Error	Adjusted R-Squared
Math	4	31.8	.05
	5	31.6	.08
	6	32.8	.04
	7	29.9	.11
	8	30.9	.02
Reading	4	28.6	.14
	5	27.1	.24
	6	27.2	.11
	7	29.0	.06
	8	30.3	.07

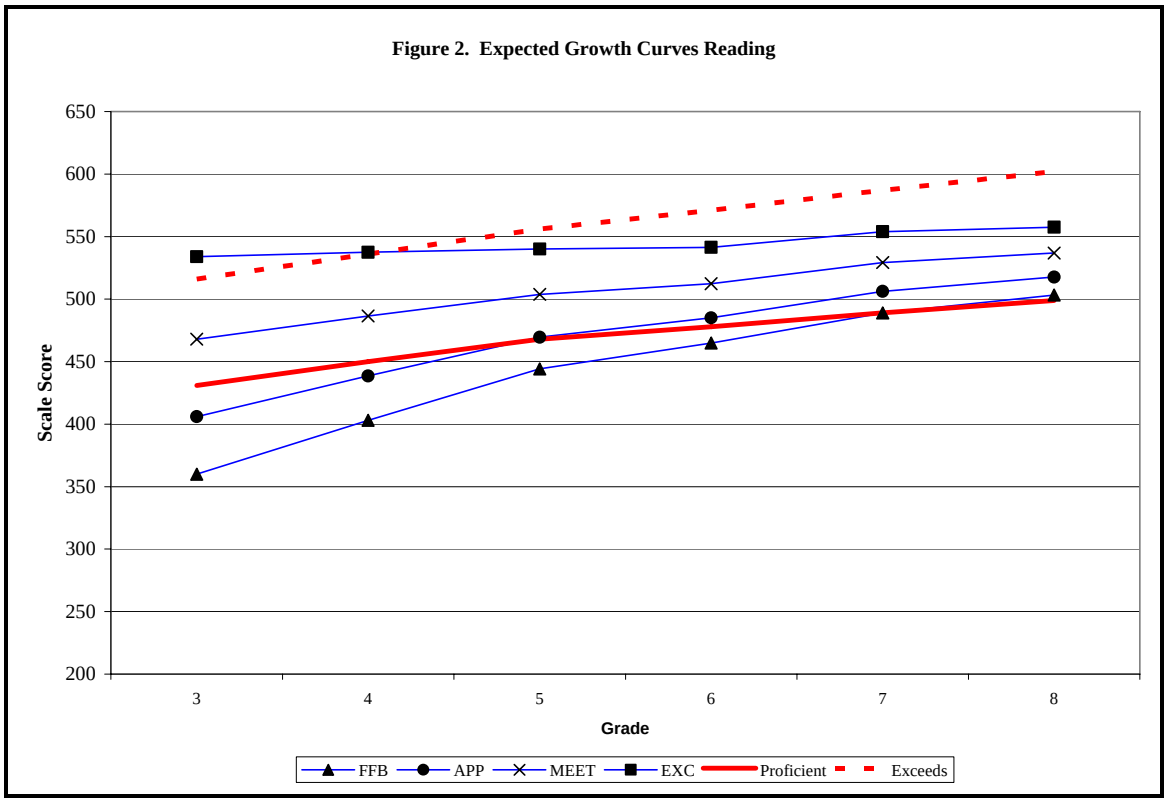
The parameter estimates for reading are given above in table 7. The parameter estimates for math were adjusted so that the growth expected would be sufficient to get the average student scoring in the falls far below category in 3rd grade to proficient by 7th grade. The adjustment was made by adding 1 to the $\hat{\alpha}$'s given from the regression analysis. Other parameter estimates were left unchanged. For example, the estimate of $\hat{\alpha}$ was 97.9308 for fourth grade math. The parameter A used to estimate expected growth is 98.9308.

The growth expectations implicit in the above model are shown in the two figures below. The figures show the expected growth curves for students earning the average score at each performance level in 3rd grade. For example, the average student score in the falls far below (FFB) performance level for 3rd grade reading was 360. The growth curves were generated by repeated iterations of the grade level equations. For example, the expected growth for a student scoring at a given level in 3rd grade was calculated using the model and added to the 3rd grade score to get an expected 4th grade score. The expected 4th grade score was then used to generate an expected growth for fifth grade. This was repeated through the grades until an expectation of where the student would wind up by 8th grade was arrived at.

A common criticism of using valued-added analysis to form growth expectations is that it sets the expectation at the state average regardless of the desirability of the average as a performance benchmark. However, the growth expectations used in the MAP analysis expect that students who start in Arizona schools in the 3rd grade will achieve proficiency in state standards by the time they reach junior high.

Figure 1. Expected Growth Curves: Math





Results of MAP Analysis

Table 9 below shows the statewide statistics for the growth index for each grade and subject. Because of the method for calculating expected growth, the average growth index is near zero for reading and -1 for math. Table 10 gives the statistics for school-level average growth indices. Comparing the two tables shows that the variation in school-level averages is much less than the variation among students. Growth indices by school, grade, and subject are reported on the Arizona Department of Education web site.

Table 9. Student Level Growth Index

Subject	Grade2006	N students	Average	S.D.	25th pctl.	75th pctl.	90th pctl.	Min	Max
Math	4	69,762	-1.0	31.8	-21	18	38	-172	201
	5	68,918	-1.0	31.6	-21	16	37	-193	198
	6	68,913	-1.0	32.8	-21	18	38	-183	194
	7	67,644	-1.0	29.8	-20	16	36	-185	194
	8	67,831	-1.0	30.9	-21	17	36	-207	249
Reading	4	69,184	0.1	28.6	-18	17	34	-150	211
	5	68,380	0.0	27.1	-17	16	34	-165	180
	6	68,722	0.0	27.2	-17	16	32	-140	202
	7	67,999	0.0	29.0	-18	17	36	-185	194
	8	68,495	0.0	30.3	-20	18	37	-185	260

Table 10. Average School Growth Index

Subject	Grade2006	N schools	Average	S.D.	25th pctl.	75th pctl.	90th pctl.	Min	Max
Math	4	906	-1.4	11.8	-9.3	6.1	13.4	-36.6	38.7
	5	893	-1.1	11.1	-8.4	5.5	13.9	-30.7	36.5
	6	741	0.0	12.6	-8.7	7.4	15.7	-38.9	48.8
	7	506	-2.1	9.9	-8.2	4.2	10.8	-31.4	30.0
	8	499	-1.7	10.6	-8.6	5.0	12.0	-50.9	28.4
Reading	4	905	0.0	7.4	-4.6	5.0	8.9	-24.8	29.9
	5	893	-0.1	7.0	-4.6	4.6	8.6	-21.4	23.4
	6	741	0.6	7.3	-4.2	5.6	9.5	-25.5	31.9
	7	506	0.2	7.2	-4.6	5.2	8.7	-24.5	25.7
	8	500	-0.7	8.3	-6.1	4.4	8.9	-30.3	30.7